

REMARKS

Applicant respectfully requests reconsideration and allowance of the application. Claims 1-26 are pending in this application.

A review of the claims indicates that:

- A) Claims 2-8, 15-17, and 21 remain in their original form.
- B) Claims 1, 9-14, 18-20, 22-26 are currently amended.
- C) No claims are previously presented.
- D) No claims are currently added.
- E) No claims are currently cancelled.

Claims 18, 19, 22, and 23 are rejected under 35 U.S.C. §112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicant regards as the invention.

Claims 18, 19, 22, and 23 are rejected under 35 U.S.C. §101, as being directed to non-statutory subject matter, and embracing and overlapping two different statutory classes of invention as set forth in 35 U.S.C. §101.

Claims 1-26 are rejected under 35 U.S.C. §102(b) as being anticipated by Altova, "User reference Manual Version 4.4, XML Spy suite 4.4", Altova Ges.m.b.h & Altova, Inc, May 24, 2002 (hereinafter "Altova").

Applicant respectfully requests reconsideration and allowance of the subject application. Claims 1-26 are pending in the application.

Information Disclosure Statement

As a preliminary matter, the Office notes that several documents were identified in IDS filings but were not found in the file. Applicant has included these documents herewith.

BEST AVAILABLE COPY

Claim Rejections under 35 U.S.C. §112 and 35 U.S.C. §101

Claims 18, 19, 22, and 23 are rejected under 35 U.S.C. §112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicant regards as the invention. Additionally, Claims 18, 19, 22, and 23 are rejected under 35 U.S.C. §101, as being directed to non-statutory subject matter, and embracing and overlapping two different statutory classes of invention as set forth in 35 U.S.C. §101.

Following telephonic consultation with the Examiner on March 1, 2006, claims 18, 19, 22, and 23 are amended to particularly point out and distinctly claim the subject matter which Applicant regards as the invention. Moreover, claims 18, 19, 22, and 23 are amended to recite elements in a method claim format. Thus amended claims 18, 19, 22, and 23 are fully enabled by the specification and are directed to statutory subject matter. Applicant respectfully requests rescission of the current rejections of claims 18, 19, 22, and 23 under 35 U.S.C. §112, second paragraph, and 35 U.S.C. §101.

Claim Rejections under 35 U.S.C. §102

Applicant wishes to thank the Examiner for the telephonic conversation on March 1 and March 6, 2006. In particular, Applicant thanks the Examiner for his indication that the claims, as amended, are neither disclosed nor shown in Altova. Applicant respectfully requests reconsideration and prompt issuance of the subject application. If any issues remain that prevent issuance of this application, the Examiner is urged to contact the undersigned attorney before issuing a subsequent Action.

Respectfully Submitted,

Dated: March 16, 2006

By: Jim Patterson

Jim Patterson
Reg. No. 52,103
(509) 324-9256 ext. 247

LEE & HAYES PLLC
Suite 500
421 W. Riverside Avenue
Spokane, Washington 99201
Telephone: 509-324-9256 x247
Facsimile: (509) 323-8979

Get More From SharePoint

A common method of Web-enabling document management systems is to make them accessible through an enterprise information portal. However, this usually requires separate portal software.

Microsoft SharePoint solves this problem by combining a document management system and document search tools with a Web portal. No separate database or Web servers are needed, and there is just one interface for users to deal with. SharePoint, however, does have its limitations.

"SharePoint can be used out of the box, but it lacks specialized high-level functions," says Gartner research director Karen Shegda. "It has limited scalability, only a very simple workflow infrastructure and limited functionality needed for vertical markets. It can be customized, however."

Enter Elite Information Systems. This Los Angeles-based company specializes in time and billing software for professional practices and consulting firms, but it recently added a robust combined document management system and Web portal built on SharePoint. The product, Elite Encompass, complements and bolsters Microsoft's technology with an enhanced interface and higher-level features required in many deployments.

Among the enhancements built into Encompass are Smart Props and Deal Room features, folder policies, offline support and file-level security. The Smart Props feature allows Encompass to load document properties from independent ODBC data-

bases. The metadata values can be loaded automatically, or users can manually enter information from pull-down menus. Deal Room lets external users securely log into the document management system. When users create documents, they can choose to keep the document private or publish it with varying levels of security, including an option for access to users outside the company firewall.

Encompass's folder policies define the physical location on a system where a document is saved. Encompass administrators can determine where files are stored according to predefined criteria, ensuring that files will be available to priority users even if the document management system goes down. This feature can also be used to force files to be saved to special kinds of storage such as DVD or magneto-optical.

Offline support lets users check out a document and work with it offline. File-level security addresses one of SharePoint's major shortcomings: It



only provides folder-level security. While documents within a folder in SharePoint have to be shared or protected as a group, Encompass lets users share or protect individual files.

Elite has also integrated time, billing and financial applications for practice management, which are available optionally. Application integration was added to bring increased functionality to the SharePoint environment. "The integration goes even deeper than at the portal's top level," says Tom Bartley, Elite's vice president of strategy. "Information freely passes between the document management system and our other applications, giving users a fully integrated environment."

At Bonne, Bridges, Mueller, O'Keefe and Nichols, a 200-employee law firm in Los Angeles, Encompass has introduced a major change in the way documents are managed.

"Prior to using Encompass, we stored files on our network using a carefully designed file naming convention," says Jeff Moffat, the firm's chief

'We had no way to share or reuse

documents. Collaboration was nonexistent.'

— JEFF MOFFAT, LOS ANGELES LAW FIRM

T
R
A
N
S
F
O
R
M
13
H
A
R
C
H
2
U
C
2

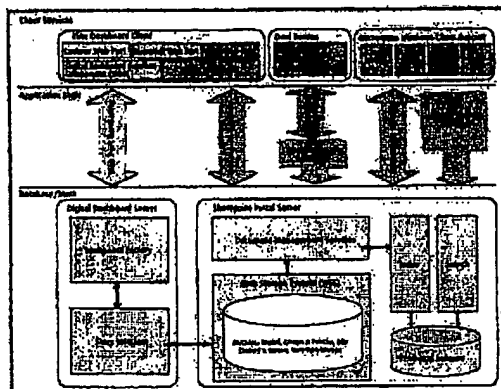
operating officer. "Most users stuck to the system fairly well, and we would lose files only occasionally. But losing any file is unacceptable."

There were other problems with the law firm's ad hoc approach. "We had no way to share or reuse documents," Moffat explains. "Collaboration on the system was almost nonexistent."

Moffat says his firm chose the software "because we had a long-standing relationship with Elite and because of Encompass's Web capabilities." The law firm was a long-time user of Elite's billing and finance management products, and Moffat says he liked the way those products could be integrated with the Digital Dashboard portal included in Encompass. Moffat adds that Smart Props will "take a lot of the pain" out of the switch to a document management environment by reducing the amount of manual data entry needed to index documents.

Input from Bonne, Bridges, Mueller, O'Keefe and Nichols during beta testing led to some important additions to Encompass. For example, the product originally supported only Microsoft Office desktop applications. But because WordPerfect is the dominant word processor among lawyers, the law firm prompted Elite to support its favored word processing tool.

Integration with productivity applications such as Microsoft Office and WordPerfect requires that Encompass still have one backward-looking software component: a dedicated client. While other file types can be searched and downloaded through Encompass, viewing and editing requires the native application. Saving back to the document management system requires Encompass thick-client software that integrates a custom Save dialog box into Office



The Elite Encompass components (in blue) complete and extend SharePoint's document management capabilities.

applications and WordPerfect.

A modern, Web-aware document management system should either use a Java applet or ActiveX control that can be dynamically downloaded to integrate with local applications or use a thin client running within a Web browser.

"Although there is a long-term trend toward thin clients in document management, thick clients are still required by many document management systems," says Shegda of Gartner.

Shegda also expressed concerns about the Microsoft technology underlying Encompass. "SharePoint is limited in its ability to scale up and has relatively primitive workflow," she says. The workflow is built on email messaging. "... Encompass will likely only be sold into small- and mid-sized companies and departments within larger companies."

At \$295 per seat, plus the cost of the required SharePoint Server (about \$72 per seat), this product's best play may be for small- to mid-size professional services firms that can also take advantage of Elite's optional but well-integrated time and billing applications. Lower-priced competitors include 80-20 Software, which has an Exchange-based document management system (\$119 per seat, plus \$6,375 to

\$9,500 per server CPU) with a new SharePoint connector.

SharePoint offers a basic management, search and collaborative infrastructure that will be a building block for more complex content and document management systems. Elite is among the first third-party vendors to release a complete product based on this platform. We expect refinements and competitive price pressures to make such offerings increasingly attractive. ●

SYNOPSIS

Vendor: Elite Information Systems, Los Angeles, www.eliteis.com

Product: Elite Encompass

Description: Document management portal built on Microsoft's SharePoint but adding features including application integration, folder policies, offline support and file-level security.

Strengths: Combines document management, Web portal, document indexing and search in a single system. Low cost relative to a best-of-breed approach involving integration.

Integrated time and billing applications for professional practices are optional.

Weaknesses: May not scale up to large numbers of users. High cost relative to simpler offerings built on SharePoint or SharePoint alone.

Price: \$295/seat for Encompass plus \$72/seat for SharePoint

New Tool Kit to Link Groove With Microsoft SharePoint. (Groove Networks SharePoint Team Services/Groove Workspace integration kit) (Product Announcement)

© COPYRIGHT 2002 Ziff Davis Media Inc.

Groove Networks Inc. this fall will make available a new tool set to connect its peer-to-peer collaboration platform with Microsoft Corp.'s SharePoint Team Services collaboration software.

The SharePoint Team Services/Groove Workspace integration kit will enable team members in different companies to share files and discussion threads and manage projects while pulling data from a SharePoint Team Services knowledge repository.

This, in effect, extends the Microsoft product beyond the firewall so that companies can work with customers and partners in a secure, data-encrypted environment. It will also allow workers using information on a SharePoint Team Services Web site to continue using the data offline.

Once they reconnect, the Groove software will synchronize the changed data with the Team Services Web site, according to officials at Groove, in Beverly, Mass., and Microsoft, in Redmond, Wash.

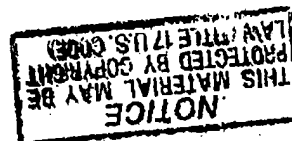
HP Services, of Palo Alto, Calif., saw the need to link a distributed, P2P collaboration application with one that provides centralized collaboration behind the firewall, such as SharePoint Team Services, said HP Services Chief Knowledge Officer Craig Samuel.

"We have the last-mile problem on the telco front, where we have lots of broadband but ... have to get it to the desktop; for me, on the collaboration side of things, Groove can solve the last-mile problem in terms of knowledge workers collaborating," Samuel said. "It lets me [connect] customers, suppliers ... and people not associated with our internal networks."

HP Services is already working to design an architecture in which teams come together in Groove Workspace and generate new knowledge, which could then be synced to SharePoint Team Services.

"Groove is not suited for thousands of people looking for something, but it is good for a couple dozen. It is the place where new knowledge is generated," Samuel said.

Officials at Groove and Microsoft said they will explore extending the tool kit to other Microsoft software. A likely candidate is Microsoft SharePoint Portal Server, which is a data repository with more sophisticated document management capabilities than Team Services.



Networking Technology - Impact and Opportunities

Simon Musgrave

Abstract

Computer networks, and in particular the Internet, have changed the way that many industries operate. The paper examines the uses of the prevailing networking technologies and their impact upon different communities. There is a more detailed examination of the promotion and dissemination of data in particular via the Data Archive. This is followed by a brief discussion of the impact of the technology upon the industry in terms of the security of data and data collection methods.

Keywords

Networks technologies, World Wide Web, distributed catalogues, data dissemination.

1. Introduction

The Internet needs little introduction to the computer literate among us. It has become the most talked about IT development of the last few years. Networking technologies have been around for many years, but it is the explosion in the use of the Internet that has grabbed the headlines. Figures for the exact growth of the Internet are hard to come by, but the general estimate is that it has been doubling every year. This growth rate has moved it from a relatively obscure UNIX based networking technology, to become the centre of a massive global information system.

This paper will review the current state of the technology and assess the impact on survey practice. Due to the author's particular expertise, it will focus in particular on the opportunities for developing new dissemination techniques for survey results.

2. Uses

The uptake of any technology is based on an interaction between the capabilities of the technology and the demands and expertise of the user community. Consequently the technology will be reviewed first, highlighting the particular strengths and weakness and this will be followed by a discussion of the uses by broad categories of users.

2.1. Technologies

There are many proprietary networking systems across the world. Most of these are related to private companies of dependent on specific hardware. Examples are DECnet for Digital VAX computers and for groupware applications, Lotus Notes now part of IBM. These applications are enormously successful within certain sectors, notably the commercial sector, and Lotus Notes in particular is the leading world-wide groupware product. What they lack is the ability to link openly with other systems. Some, however, may argue that this is a strength. A plethora of bridges and routers have been developed to handle the connections, but analogous to transfers between survey packages, many problems of compatibility can occur and some functionality is lost in the process.

Over the last few years these proprietary systems have been overtaken by the de facto Internet (or TCP/IP) protocols. There have also been de jure standards which are discussed very briefly.

2.1.1 Internet

The Internet suite of protocols was developed from the needs of the US military to develop a fault tolerant networking system. As such the Internet is a connectionless network. In other words it operates more like the Post Office than the Telephone, where each packet has an address and can take several routes to reach a destination. If the system is busy, it will take longer to reach that destination. This is the converse to the telephone in which all calls are carried at the same rate, and congestion means that there will not be enough lines to carry the volume of traffic. In order to provide background to the impact and opportunities discussed later, the main application protocols¹ are introduced briefly.

- Telnet is the service which provides a terminal connection on a remote computer. This is a well established protocol, dating back to the earliest use of networking. Nevertheless most users have experienced problems with key mapping at some stage or another.
- FTP stands for File Transfer Protocol and is concerned with the transfer of files between locations. It is another well established service and yet becoming more important as an integrated part of more innovative services, such as the World Wide Web.
- Z39.50 is the protocol for the interrogation of remote databases. The most popular example is the WAIS database system, which has several releases. It is developing into a powerful tool for fielded searches across different sites. Additionally it is very effective at handling different data types. An example of its use for catalogues of survey data will be discussed later.
- HTTP is the protocol for World Wide Web services. These services are often seen as synonymous with the Internet itself, but in fact are relatively new on the scene, having been developed in CERN in the early 90s. It is the emergence of browsers such as Mosaic, in 1993, and Netscape, in 1994, with their sophisticated ability to integrate multimedia parts, that has led to the enormous growth in usage.
- Gopher. An earlier service to the World Wide Web was the gopher system, developed at the University of Minnesota. This was an effective information system that has many of the facilities of the WWW, in particular the handling of different file types. This led to a rapid growth in Gopher usage in the early 90s. Its use has now been eclipsed by the WWW.
- SMTP is the electronic mail protocol. Again electronic mail has a long history in networking, being one of the first applications. It is now very widespread but SMTP is only one of several competing protocols. Nevertheless the simple addresses, e.g. s.w.musgrave@essex.ac.uk have become widely known and used by commercial, academic and government users.

2.1.2 OSI

An alternative set of protocols is provided by the OSI standard. OSI was set up through a variety of committees, mainly in the 80s. X.400 is the email protocol, although this went through various flavours. FTAM is the file transfer protocol. One useful standard not provided by the Internet is the X.500 directory service. This is a database system for addresses of networked organisations. This standard has been included in some Internet applications.

¹ These are the standards for the envelope contents, the transport protocol (TCP/IP) is the envelope.

2.2. Communities

The uptake of computing technology has been very different among separate parts of society. These communities of users have taken up these technologies at very different rates. Academia have been leaders and the least franchised are the general public, although this is changing fast.

2.2.1 Academia

As with many new IT developments, academic use has been at the forefront of the evolution of the Internet. It has been tested, hacked and enhanced by many research institutes worldwide. As with UNIX and graphical interfaces, many of the rough edges have been smoothed before it reaches the wider community. This has meant somewhat patchy development. As an example, given the academic users' infrequent need for secure transaction, developments on security aspects of the Internet have had to wait until it was used by the commercial sector.

Education is a core business for academia and for that reason many breathtaking applications have been developed that take advantage of the hypertext capabilities of the WWW. Many of these were developed in earlier days on Apple's HyperCard or similar products. Research is the other core activity of academia and thus explains the proliferation of major resources on the Internet. Examples related to survey research include the data catalogues in many countries as well as on-line analysis systems such as MIDAS².

2.2.2 Public Sector

The public sector in the UK is embracing Internet developments rapidly after a delayed start. The US is well known for being a leading supplier of public information on the Internet. A good example is the US Census bureau with its large on-line database of demographic and economic statistics. Canada, Australia and Singapore are also prolific publishers of information.

Local Government has been slower to take advantage of the Internet. Up to now it has not been cost effective to set up services of limited use to this community, typically local business and the public. However with initiatives to link schools and libraries, this may change.

2.2.3 Private Sector

The private sector have made extensive use of networking, in particular large companies with extensive office networks, or those companies that deal with information directly, for example banks. Networking has been based upon proprietary standards until recently. Usually this was quite sufficient as there was little need to connect to external services. However, over the last couple of years most major firms have established external links as demonstrated by the increase in the number which have set-up their own web page. For many this is simply ensuring that they have an electronic brochure. It is difficult to tell how many serious users are looking at the pages, but unexpected results do arise. For example, some small industries in particular have found it a useful way of exploring new markets, selling in distant countries or finding new business partners. The development of security for transactions means the electronic commerce is expected to grow rapidly to become major business by the end of the century. The cost of electronic data exchange is only about 10-30% of the traditional EDI³. All of these developments have implications for survey business. The distribution of

² MIDAS stands for Manchester Information Datasets and Associated Services

³ Survey by Mastercard International, November 1995, quoted FT, April 3 1996

results of a survey is the most obvious application, but there is likely to be a significant increase in the number of surveys conducted by collecting information via the Internet.

2.2.4 General Public

In the UK, it is the general public who are least directly affected by the Internet at present. Whilst some home users have bought computers and paid for Internet access (typically about £10 a month plus phone bills), these are, typically, the young, male, affluent and well educated!

However, there are important developments which may help to spread the impact of the technology. The first is schools which are increasingly networked. This gives young people a chance to become acquainted with the technology. The second is public libraries which take their role of public information access very seriously and so are frequently setting up Internet access points (though they tend to be very busy and thus can only provide limited access). The third is the emergence of cyber-cafes where customers can obtain free access to the Internet.

3. Opportunities

In the discussion above, I have identified the WWW as the key application behind the growth in network use and the driving force in the development of new applications. The pages of the national and computing press are full of discussion of the products and strategies of the leading players in the market, such as Microsoft, Netscape, Sun, Oracle and IBM with its particular interest in Lotus Notes. In the early days WWW use was criticised as being the preserve of the enthusiasts, and yet recently this has become a mainstream application as most major companies have ensured that they have a presence on the WWW. Co-operative work and specifically the Intranet (the use of Internet products and standards within rather than between organisations) is the latest 'hot' topic. I now turn to its use within the survey community now that the technology is well established. There are many aspects to the potential impact of the technology. I will start with the general applications and for these the UK Data Archive will be used as an example. I will also refer to NESSTAR⁴, a project with EU funding under the IV framework programme. This has been obtained to determine the potential for the further development of the integrated aspects of the applications described below. Wider issues of security and data mis-use will be discussed in the subsequent section as well as the likely impact upon the industry.

3.1. Brochures

The majority of WWW sites (estimated at 95%) are simply electronic brochures. Nevertheless they are now seen as an essential feature of any major business. They are an extension of advertising and publicity and one only has to look at some of the more colourful displays to realise that serious money has been committed to their creation. It gives a new opening for graphic artists and it is expected that as competition in cyberspace grows, so the ability and skills to attract attention with powerful graphics and other multi-media will be at a premium.

Official producers of data have not been slow to make their presence felt on the Internet, and the major data collections are advertised via this medium. Few UK sites offer access to the data themselves however. Effective use of these medium requires a combining of the skills of statisticians, publishing or marketing experts and technical people.

⁴ NESSTAR, Networked European Social Science Tools And Resources is a consortium of the Norwegian, Danish and UK Data Archive's to develop network based tools.

3.2. Catalogues

3.2.1 Remote Access

A second common use of networking is to provide a front end to catalogues. Such catalogues are normally information which is in the public domain rather than restricted and their use is encouraged. Front ends were an early application of networking technology and historically have been based on VT100 type interfaces. More recently WWW front ends have been provided to these databases. However, rather than the systems being based on the common hypertext HTML structure of normal WWW pages, they usually utilise databases tools to link into more powerful database systems. For the Data Archive, this means that the catalogue of data holdings can be searched using a form mounted on the WWW which links into the SQL database.

3.2.2 Integrated

A major advantage of network tools is the ability to create "virtual catalogues". What can appear to the user as one catalogue can, in fact, be made up of many different catalogues. An example of this is the integrated catalogue of the European Data Archives. This catalogue consists of a WWW based front end, which is then linked to a number of separate catalogues which can be selected by the user from a master list. See picture below.

The screenshot shows a web-based search interface for the CES S D A Integrated Data Catalogue. The interface is divided into several sections:

- Archives to search:** A list of archives with checkboxes for selection:
 - ☐ BDSP, France
 - ☐ DDA, Denmark
 - ☐ DA (UK), UK
 - ☒ NSD, Norway
 - ☐ SSD, Sweden
 - ☐ STAR, Netherlands
 - ☐ TARKI, Hungary
 - ☐ ZA, Germany
- Full text search:** A single text input field with a search button.
- Fielded search:** Multiple input fields for specific search criteria:
 - Title:
 - Names:
 - Contents:
 - Start year: [] End year: []
 - Geographical focus:
- Query options:** Radio buttons for "Connect fields with: AND OR", "Verbose list: YES NO", and a text box for "Max. number of hits: 40".

At the bottom, there are two buttons: "Start search" and "Reset query".

Copyright © Norwegian Social Science Data Services, 1996
Please email any comments to webmaster@nsd.uib.no

On issuing a search, a Z39.50 protocol is used to send search instructions to the WAIS databases. These results are then sorted and displayed via the WWW. The user does not need to know where the search being sent (although bandwidth problems may slow this down). Also or similarly the data providers do not need to merge all the separate services into one master catalogue, the network does it automatically. This catalogue front end is the starting point for many of the NESSTAR tools for integrated data browsing and access.

3.2.3 Search Tools

As well as the applications built specifically to search particular applications, it is possible to use a number of WWW search tools as well. The growing number of WWW search engines include Yahoo, Inktomi, Alta Vista, Open Text and Excite⁵. It will take a while for a clear market leader to emerge, but what all these services provide is the ability to gather large amounts of data into indexed databases ready for rapid searching.

The availability of these tools makes it feasible to find any information about any subject that is on the network. It is an attempt to bring some order to the assumed anarchy of the WWW. However these do not provide structured search environments, such as those provided by BIRON, the Data Archives' on-line catalogue⁶.

3.3. Browsing Systems

As well as searching catalogues, the user may want to browse and search documentation and data itself. These searches may well provide a preliminary taster to the user, before they decide whether or not to order a dataset, or a full service in itself.

3.3.1 Text

A first step towards obtaining and using a dataset is to examine accompanying documentation. The documentation may have two main parts. First, there is the list of variables and their labels, sometimes known as the codebook. These can be searched to determine whether the survey data contains the information required. The second type of documentation will consist of a wide variety of documents including a description of the background to the survey, the interviewers instructions and questionnaire except for some surveys conducted using computer generated questionnaires (CAPI⁷). It can include links to the variable list (or question text).

As with catalogues it is possible to link to powerful text searching software behind the WWW pages. This allows an organisation to deliver effective tools for text analysis without having to provide a database application direct to the user.

On a more practical level, the provision of documentation across the network, means that the secondary analyst does not need to purchase paper documentation when it can be read on the network. This can be particularly important for the user of a series of datasets who may be interested in just a few variables per year.

3.3.2 Data

As for text, data can be accessed using powerful search and browse facilities provided across the network. This has several benefits. Firstly, data can be sampled before purchase or acquisition. This is particularly important for large and complex datasets. Secondly, immediate results can be obtained for simple analyses. This is much quicker than downloading a complete survey for analysis.

⁵ Glyn Moody, New Scientist, 6th April 1996, pp37-40

⁶ Available at dawww.essex.ac.uk

⁷ The use of CAPI means that the secondary analyst may be able to view the structure of the questionnaire. Whilst such techniques are an advantage for the data collector, it does make it more difficult for the secondary analyst to follow the structure of the questionnaire.

There are a growing number of specialist browsing systems that can be downloaded via the WWW⁸. These require the data to be delivered in a particular format, but they do facilitate effective browsing and simple visualisation to take place quickly and easily.

3.3.3 Visualisation

An application area that is still in its infancy on the network is the use of visualisation. This has enormous potential to aid in the rapid browsing of survey data. There are some novel applications, particularly those utilising Geographic Information Systems⁹. However the effective use of graphics for the display and analysis of surveys is still focused on stand alone applications. These applications can be linked to WWW pages, but only with considerable effort and so we can expect more developments in this area, probably by utilising the JAVA programming language. The meeting of the requirements, as outlined by Tufte¹⁰, and the technology is expected to produce some exciting techniques for the display of survey data. This is one of the topics being researched within the NESSTAR project which is aiming to exploit simple graphics for the display on information on the network.

3.4. Dissemination Systems

The network is particularly good for the delivery of data and documentation. File transfer (FTP on the Internet) is one of the most well established networking applications. Its use has grown steadily for the dissemination of survey data and results. Nevertheless it is the Data Archive's experience that the use of a CD ROM for large datasets is often more efficient than the use the network. This is because a CD ROM, with its capacity of 650 Mbytes, is often larger than the space available to a user for FTPing a large file.

FTP has received a boost via the WWW as the latter utilises the FTP protocols for the delivery of files. As a result it has become integrated with many major web sites for the delivery of updates, trial datasets and similar files.

3.4.1 Subsetting and Conversion

An application area which the Data Archive is currently developing is the use of user driven subsetting and conversion on the network. Previously the Data Archive asked its users to indicate on a list those variables from a selected dataset which they wanted. These were typically loaded into a SIR¹¹ retrieval which then output the data in one of several popular formats. However, the Data Archive is in the process of placing all the variable lists for the large datasets on the WWW, linked to both the catalogue and data access forms. This allows the users to create their own subsets and gives them the option of choosing one of several main formats. Immediate FTP would be attractive for those users with small subsets, others are more likely to require a CD ROM to be written for them.

⁸ Examples include Ivision (Ivation Systems, <http://www.ivation.com/>) and Navidata from the Office of National Statistics, UK.

⁹ Examples include the KINDS project at Manchester Metropolitan, MIDAS and Salford (<http://cs6400.mcc.ac.uk/kinds/>) and the Argus Project project at Leicester (<http://www.geog.le.ac.uk/argus/index.html>).

¹⁰ 'Envisioning Information' E. Tufte, Graphics Press, 1991

¹¹ SIR stands for Scientific Information Retrieval and is the main data management package used for large government datasets.

3.4.2 Data Merging

Merging can happen at several stages in the survey process but there are both technical and conceptual problems to overcome. There are advantages in harmonising datasets at the collection stage, by the use of consistent sampling frames and identical questions and question definitions. If this can be achieved, then a pseudo panel data set can be created. Such merging can take place across waves within a survey. However, merging across surveys is more complicated and weighting factors may have to be applied before carrying out this kind of meta-analysis. What is important about the network is that it facilitates the access to a variety of data sources that can be combined at the point of analysis.

3.4.3 Integration

As discussed in the section on data merging, there are a growing number of tools that facilitate integration of data resources. One the one level there are efforts to standardise data formats and in particular the codebook descriptions¹². On another level there is a growing movement towards the integrated desktop. This would allow a variety of information sources and analysis tools to be combined at the desktop, allowing easy movement from one to another. For example the user may have a WWW link to a documentation system, another link to a large dataset, a local graphics system and finally a word processor to bring together a report.

The Data Archive, through its programme of digitisation of documentation, will be making the question text of the variables of most surveys available on-line. The effect of this will be to enable a user to identify all questions across many surveys that have the same themes. This takes the user to a level of information that a catalogue cannot provide, however well indexed.

4. Impact

4.1. Security

The security implications of the network and WWW in particular is a subject of great interest and one which has generated a great deal of debate. Concerns about hackers, viruses and data theft have grown. Data on a standalone machine have many risks, such as corruption and accidental deletion, but by placing them on a networked machine, another set of risks become important. Various techniques have been, or are being, developed to address these issues.

4.1.1 Firewalls

Firewalls are computer systems that ensure that only authorised information is allowed to enter or leave an organisation. They are an essential part of any organisation's access to the Internet. Various tools have been developed to test out network security and the demands of commercial transactions are likely to lead to tighter security measures being created.

4.1.2 Tattoos

Of major concern to some data producers is the necessity to ensure that the source of data is acknowledged and that any payments due are made. Dissemination of data via networked systems can utilise techniques to stamp data with the owner's logo or some other means of identification. This is known as tattooing. In the music industry this has taken the form of marking each file with an

¹² Examples include the Triple S consortium which has published a standard and the IASSIST (International Association for Social Science Information, Support and Technology) DTD (document type definition) committee, led by the ICPSR (Inter Campus Consortium for Political and Social Research) at the University of Michigan, USA.

inaudible code. This code can be automatically detected when broadcast. This can have implications for data collection, as discussed below.

For survey data the techniques employed in the music and video industry are hard to apply to straightforward ASCII text. They could be applied to system files, but these would then become system specific. Nevertheless it is a possibility to lock data into a read only system. The only way of getting hold of it would be via a cut and paste operation and then it would probably be cheaper and easier to pay for it.

4.1.3 Access

An alternative to the above techniques is to distribute locked copies of data. This has been developed in the COPICAT programme¹³. In this technique an encryption algorithm is run which can only be unlocked by a modification to the file system of the users computer. The encrypted file can be distributed across the network, but only authorised users can access it. It is better suited to text than data, but the technique does have some possibilities.

4.2. Future of Industry

4.2.1 Data Collection

There have been significant changes in the way surveys have been collected over the last few years. The most important of these are the use of computer assisted methods such as CAPI and CATI. These have improved the turnaround of results and enabled better quality data to be collected in part because of the immediacy of validation.

At the moment the use of the Internet is similar to early use of the phone for data collection. It was the rich and those with particular needs who were the early innovators. For a long time surveys were not conducted by telephone because of the obvious biases that would be introduced. However for many surveys this is no longer a constraint. Similarly it can be expected that most households will have access to advanced electronic information systems in the next ten years. These could easily be used to collect data 100% data in areas that were previously the domain of the survey. An analysis of telephone call patterns is an example. It is straightforward to log all the basic information about telephone calls and produce extensive reports, without having to rely on surveys at all.

More pragmatically, many data collectors of commercial data are relying on disks being sent out to users for completion on the computer. Wherever possible these are linked to administrative systems. It is straightforward for these systems to be developed on the WWW and many users have indicated such a preference. In a few years it can be expected that most data will be collected via forms on the WWW or via some sort of automatic recording of information, such as supermarket loyalty cards.

4.2.2 Usage Recording

It is not just in the collection of data that these techniques are applicable. Many suppliers of data want to know who is using the data that has been collected. If techniques of fingerprinting data can be used to record usage, then the data suppliers will be able to obtain a better return on their investment and data resource centres like the Data Archive will be able to audit usage. Both would lead to more data in areas of demand.

¹³ The COPICAT programme is an ESPRIT project on Copyright Ownership Protection in Computer Assisted Training), ref.: www.mari.co.uk/~copi/copicat

5. Conclusions

The following table summarise the main advantages of using networking technology in the survey business, as opposed to the traditional methods, whether for collection, analysis or any other aspect of the work.

	<i>Traditional Methods</i>	<i>Network Use</i>
Data Collection	Expertise already established	Improved Accuracy Rapid and perhaps cheaper data acquisition
Data/documentation browsing	Powerful local tools	Rapid access Cost effective
Data analysis	Familiar powerful tools	Access to central resources
Dissemination	More controlled	Rapid delivery Flexible delivery
Promotion	Targeted to known users and mailing lists	Widespread Unexpected results

The outlook for traditional survey collection analysis and dissemination is undergoing massive changes. At the same time the requirements for skills to analyse these growing mountains of data are growing. This has spurred the industry to develop new techniques in data warehousing and mining. There are now well established lines of work building on the analytic power available on modern machines. What remains to be fully exploited is the network to collect, catalogue, view, promote and disseminate this data. Software will need to be able to handle larger amounts of data and to give results in more immediately understandable ways, but perhaps more than anything, it must be able to link with the growing network resources.

About the Author

Simon Musgrave has been the Technical Director of the UK Data Archive since 1993. Previously he was User Services Manager of Aston University Computing Service and before that a Microcomputer Consultant in the Computing Service and Research Associate in the Development Economics Research Centre at the University of Warwick. Currently his main interests are in data management and Internet developments.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☒ BLACK BORDERS
- ☒ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☒ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.